Prefixational Object Perception in Scenes:
Objects Popping Out of Schemas

Peter De Graef
University of Leuven, Belgium

Semantic influences of context on the ease of object identification in real-world scenes are commonly accepted, but when eye movements are taken into account the unanimity dwindles. The question is whether object-in-scene semantics only come into play when the object is foveated or already have an impact during extrafoveal, prefixational object processing, and if so, whether semantic consistency or inconsistency would enhance extrafoveal processing. A theoretical framework (mismatch theory) is borrowed from reading and word recognition to support the hypothesis that both consistency and inconsistency may facilitate extrafoveal processing. Two earlier studies of context-sensitive object identification in scenes are reanalyzed to provide an initial test of the validity of the theoretical framework. Analysis of context effects on gaze shift frequency, gaze shift destination and gaze shift latencies suggests that in the earliest stages of scene exploration scene-inconsistent objects are more salient saccade targets. However, this does not appear to be a popout phenomenon based on attentional capture by schema-inconsistent objects, but rather reflects a smaller useful field of view for such objects. If attention is selectively captured at the onset of scene exploration, it appears to be by schema-consistent rather than inconsistent objects.

When comparing reviews of eye-movement research on real-world scene perception one finds unanimity on one conclusion: Objects fixated in a scene they are likely to appear in are easier to identify than the same objects fixated in an implausible context (Boyce & Pollatsek, 1992a; De Graef, 1992; Henderson, 1992a; Rayner & Pollatsek, 1992). This unanimity is hardly surprising given the repeated finding that plausible objects exhibit shorter first fixation durations (De Graef, Christiaens, & d'Ydewalle, 1990), shorter gaze durations (Antes & Penland, 1981; Friedman, 1979; Loftus & Mackworth, 1978), and shorter naming latencies (Boyce & Pollatsek, 1992b) than implausible objects. As will be documented below, however, less unanimity exists with respect to plausibility effects on the identifiability of objects that are not being fixated. Logically, this implies extrafoveally located objects that haven't been fixated yet (i.e., prefixational) or objects the eye has already left behind (i.e., postfixational). In the remainder of the discussion, I will focus on the former category because, to the best of my knowledge, no empirical data are available for the latter.

Prefixational Perceptibility:
Popout versus Schema-Driven?

Empirical data on this issue are not abundant but this has not prevented the emergence of strong claims about the impact of context on the extrafoveal perceptibility of prefixational objects. A recent example is Christie and Klein's (1995) reference to "the well-established phenomenon of unexpected items popping out in natural scenes (Loftus & Mackworth, 1978)" (p. 550). In citing Loftus and Mackworth, Christie and Klein follow a long tradition in which this paper is cited to support the claim that objects that do not belong in a given real-world environment stand out and
immediately and inexorably draw the viewer’s attention and gaze (e.g., Johnston, Hawley, Plewe, Elliott, & DeWitt, 1990; Marks, McFalls, & Hopkinson, 1992; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). Among researchers of scene perception, however, this claim has received little support. Henderson and Hollingworth (this volume) provide an exhaustive overview of the criticisms, so I will limit myself to three points pertinent to the present chapter.

First, the Loftus and Mackworth finding that an implausible object in a scene is fixated earlier on in scene exploration than a plausible object at the same location in the same scene has frequently been criticized for a possible confound between implausibility of an object and its visual dissimilarity from the rest of the scene (e.g., Rayner & Pollatsek, 1992; Henderson, 1992a). Second, subsequent eye-movement studies failed to find a fixation precedence for implausible over plausible objects either when they are presented together, at different positions in the same scene (Friedman & Liebelt, 1981) or when presented separately at the same position in separate copies of the same scene (De Graef, Christiaens, & d’Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1997). Finally, when extrafoveal target objects need to be discriminated in briefly flashed scenes, performance is consistently better for plausible objects which appears to indicate the opposite of implausible popout (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989; De Graef & d’Ydewalle, 1995).

Based on the ambiguous evidence for attention/gaze capture and on the superior extrafoveal discriminability of plausible objects, several authors have suggested that object plausibility increases the useful field of view (Antes & Penland, 1981; Biederman, Mezzanotte, Rabinowitz, Francolini, & Plude, 1981; Friedman, 1979). In other words, the perceptibility of prefixational objects in real-world scenes is enhanced when they belong in the scene. Following Friedman (1979), the main mechanism underlying this enhancement is the activation of a scene-specific schema stipulating the gross features of the objects likely to be present in the scene. For instance, activation of a kitchen schema could sensitize the visual system towards detection of a refrigerator, defined as a large, upright, shiny and white, brick-shaped object. Wolfe and Bennett (1997) recently argued that features of this type (with the possible exception of shape) can be preattentively segmented from a scene and correctly assigned to an object. According to Friedman (1979), the mere detection of such a schema-specified bundle of features may in itself be sufficient to trigger object identification. Alternatively, the feature-bundle may constitute a salient target for an attention shift followed by enhanced data-driven processing at the attended location (Antes & Kristjanson, 1993; Antes & Penland, 1981). Either way, prefixational perceptibility will be greater than that of objects which are not part of the activated schema and whose diagnostic features need to be extracted and bound in a foveal analysis of high-resolution detail.

From the above, it should be clear that two quite different views are being advocated with respect to the context-sensitivity of prefixational object perception in real-world scenes: Schema-driven facilitation of context-consistent objects on the one hand, and perceptual popout of context-inconsistent objects on the other hand. Surprisingly, these two views are primarily held as convictions and no systematic attempts have been made to either reconcile them or unequivocally decide between them. The present chapter does not pretend to put an end to this state of affairs. Rather, I would like to set the stage for future experimental investigation of this neglected issue in scene research. Specifically, I will indicate some potentially relevant hypotheses developed in research on word recognition and reading. As a preliminary test of their applicability to prefixational object perception, they will be pitted against previously unanalyzed eye-movement data which I collected in an ongoing series of studies aimed
at establishing the boundary conditions under which object fixations in scenes exhibit context effects.

A Mismatch Theory of Prefixational Perceptibility: Popout and Schema-Driven

In an extended series of experiments, Johnston and colleagues presented viewers with brief flashes (33-400 ms) of four-item arrays containing either words (Johnston et al. 1990; Johnston, Hawley, & Farnham, 1993) or nonsense symbol-strings (Hawley, Johnston, & Farnham, 1994). Following a mask, viewers were given a probe word or string and were asked to indicate where it had been located in the array. Since the arrays subtended about 6° x 4° and exposure durations allowed for one eye movement at most, perceptual processing in this task had to be primarily prefixational which explains the potential relevance of these studies to our current topic of discussion.

Analysis of the percentage of correct localizations showed three basic effects. First, the baseline effect, that is, superior localization in all-familiar arrays -which repeatedly combined the same items- than in all-novel arrays -which consisted of previously unseen items-. Second, novel popout, defined by better localization of single novel objects in otherwise familiar arrays (i.e., one-novel arrays) than for the same objects in all-novel arrays. Third, familiar sink-in, reflected in lower accuracy for familiar objects in a one-novel array than in all-familiar arrays.

In order to coherently explain these findings, Johnston and Hawley (1994) argue that the effects show that schema-driven perception and popout need not be mutually exclusive. In fact, they consider it to be an essential characteristic of the adaptive mind to refrain from superfluous data-driven processing of predictable input and at the same time remain alert for novel input. According to Johnston and Hawley's mismatch theory, this is achieved by an automatic learning process through which repeated co-occurrence of the same items in an array fuses the items into a unitized array. This process has three consequences for future encounters with the unitized arrays. Specifically, during the initial glance parallel processing of array items will very rapidly produce 1) increased concept-driven processing of items that match expectations, 2) suppressed data-driven processing of the same expected items, and 3) enhanced data-driven processing of items that do not match expectations.

At first sight, this computational model of perceptual facilitation and inhibition in experimentally learned arrays covers a great deal of the available data for experientially learned scenes. First, increased concept-driven processing of expected inputs explains why plausible objects require shorter fixation and naming times in extended scene viewing and exhibit superior prefixational detection and forced choice recognition in briefly flashed scenes. Moreover, the limitation of the hypothesized processing effects to unitized arrays would explain why no plausible advantage is found when object arrays are used instead of full scenes (Antes & Penland, 1981; Antes, Penland, & Metzger, 1981; Biederman, Blickle, Teitelbaum, & Klatsky, 1988; De Graef, 1990; Henderson, Pollatsk, & Rayner, 1987). Second, suppressed data-driven processing of expected items would explain inferior memory for the presence, location and featural details of plausible objects in real-world scenes (Friedman, 1979; Henderson & Hollingworth, 1997; Pezdek et al., 1989). Finally, enhanced data-driven processing of unexpected items could account for a fixation precedence for single implausible objects in scenes (Loftus & Mackworth, 1978) under the assumption that attention/gaze shifts to a particular location are a consequence rather than the cause of heightened input processing at that location (Dark, Vochatzer, & VanVoorhis, 1996; Hoffmann, 1987; Johnston & Hawley, 1994). The disappearance of the implausible fixation precedence in non-unitized arrays of episodically related objects (Antes & Penland, 1981; De Graef, 1990) is also
consistent with mismatch theory.

In spite of its intuitive appeal as a comprehensive account of plausibility effects in scene perception, mismatch theory still does not seem to resolve the controversy surrounding prefixational object perceptibility: The theory appears to predict earlier fixation and superior extrafoveal discrimination of implausible objects, and neither prediction is confirmed. However, the solution to this problem may lie in a recent exchange about the reliability of novel popout in word arrays. Specifically, Christie and Klein (1996) criticized the work by Johnston and colleagues for failing to establish replicable and unconfounded within-array novel popout, that is, better localization of the novel than of the familiar objects in one-novel arrays. Johnston and Schwarting (1996) conceded that within-array novel popout is a somewhat elusive effect because its emergence depends on the relative magnitude of all three perceptual processing effects defined above. Specifically, if concept-driven facilitation of expected items is too large to be outweighed by suppression of expected-input processing and enhancement of unexpected-input processing, there will still be a perceptibility advantage for the expected inputs in one-novel arrays. Hence, Johnston and Schwarting argue, between-array novel popout is a better diagnostic. If unexpected inputs in one-novel arrays do draw the focus of data-driven processing, then perceptibility of the novel singleton should be superior to that of the same object in an all-novel array.

When applied to scene perception, the within-array vs. between-array distinction may resolve the schema vs. popout debate. Specifically, it suggests that only in scenes with a low to moderate conceptual facilitation of the plausible objects in them, one may be able to observe fixation precedence and superior extrafoveal discrimination of the implausible intruder relative to its plausible companions in the scene. In scenes with a strong conceptual facilitation of plausible objects, a perceptual advantage for the implausible object may disappear or reverse: Either the extrafoveal detection of schema-specified object features allows superior prefixational identification of plausible objects, or it more effectively segments the plausible object from its background and thus provides a more salient target for an attention shift. Hence, the repeated failures to replicate the Loftus and Mackworth (1978) finding of an implausible-object fixation precedence need not rule out that object implausibility increases attentional saliency. Instead, the easier segmentation of plausible objects from the densely packed scenes used in the later studies (De Graef et al, 1990; Friedman & Liebelt, 1981; Henderson et al., 1997) may have swamped an implausible-object popout effect which did surface in the sparsely populated scenes that were used by Loftus and Mackworth (1978).

Testing a Mismatch Theory of Prefixational Object Perceptibility

Future tests of the mismatch account of scene perception will require two levels of research. A first level is to establish the three between-array effects: 1) the baseline effect, that is, better perceptibility of plausible objects in a scene populated with plausible companion objects (all-plausible scene) than in a scene with only implausible and episodically unrelated companions (all-implausible scene), 2) familiar sink-in, that is, better perceptibility of plausible objects in an all-plausible scene than in scene containing an implausible singleton in a company of plausible objects (one-implausible scene), and 3) novel popout, that is, better perceptibility of the implausible singleton in a one-implausible scene than in all-implausible scene.

A second level at which the theory needs to be tested is that of the perceptual processing hypotheses advanced by Johnston and colleagues: Do plausible objects exhibit enhanced concept-driven as well as suppressed data-driven processing and is enhancement of data-driven processing limited to implausible objects? It is at this level that I want to introduce some data which appear to question
Christie and Klein (1995) presented subjects with arrays of two extrafoveally located letter strings, one a regular word, the other an unpronounceable nonword. After a brief exposure of 100, 200, or 400 ms, one of the strings was shifted up or down and the subject was to report the direction of the shift as quickly and accurately as possible. The important finding in this study was that shift detection for the regular words was superior to that for non-words, but only for the 100 and 200 ms exposures. Christie and Klein conclude that this shows an initial period of enhanced data-driven processing of familiar items caused by a rapid capture of attention. Further support for processing precedence for familiar items can be found in the work of Hoffmann (1987), who demonstrated that in arrays of extrafoveally located objects, the pictures with fast basic-level categorization times (approx. 485 ms) invariably drew more attention than the pictures with slower categorization times (approx. 510 ms). Similarly, Dark et al. (1996) reported that in 100 ms exposures of two-word arrays, attention was captured by the extrafoveal word that was semantically related to a central prime presented just prior to array-exposure.

These data indicate that the prevailing activation level of a lexical or conceptual node strongly influences the perceptual system's reactivity to any evidence in the visual field for that node's referent. Applied to the domain of prefixational object perceptibility this suggests that not the implausible but the plausible objects should initially capture attention and data-driven processing because their conceptual representations receive a surplus activation from context.

Initial Attention Capture: Evidence from the Wiggle-Paradigm

In order to determine to what extent scene-context effects on object identification are caused by inter-object priming and/or object-in-scene plausibility, De Graef, De Troy, and d'Ydewalle (1992) conducted a study in which viewers were asked to search for non-objects embedded in a black-on-white line drawing of a real-world scene. While participants could freely explore the scene, we did attempt to steer their first saccade from a designated prime object to a designated target object by means of a technique borrowed from Boyce and Pollatsek (1992b). As shown in Figure 1, each trial started with a fixation cross on a blank CRT-screen. When participants fixated the cross for at least 200 ms, as registered by a Generation 5.5 dual-Purkinje-image eye tracker, the stimulus control program automatically initiated an 8-s exposure of a scene. Stimuli always contained an object at the former location of the fixation cross (the prime object) as well as a peripherally located object which rapidly moved up and down after 160 ms of scene exposure (the target object). By introducing this target wiggle during fixation of the prime we hoped to elicit an automatic orienting response from prime to target. While we were primarily interested in target fixation parameters as a function of the target's relation to the background and the prime, this study now provides us with some data on prefixational target perceptibility.

Specifically, because the wiggle started at 160 ms delay from scene onset we can test the notion of an initial processing capture by the conceptually activated scene-consistent objects. Following Christie and Klein's (1995) logic we can assume that if plausible objects initially capture attention, an early wiggle of a plausible target should be easier to detect. Because our participants weren't informed about the wiggle an implicit measure of wiggle detection needed to be computed. For this purpose, I selected all trials on which the wiggle occurred during the first scene fixation, which in this experiment always fell on the prime object. On these trials, two groups of indexes of ocular reactivity were computed. The first group pertained to the saliency of the wiggled object as a saccade target and included three measures: 1) the proportion of trials on which the viewer's gaze moved directly from the prime to the target...
Table 1

Effects of Wiggled Target’s Plausibility on Direct Target Hits, Target Skips, and Number of Intervening Fixations preceding Delayed Target Hits (Wiggle SOA = 160 ms)

<table>
<thead>
<tr>
<th>Target</th>
<th>Plausible</th>
<th>Implausible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct target hits (%)</td>
<td>35.0</td>
<td>35.8</td>
</tr>
<tr>
<td>Target skips (%)</td>
<td>17.5</td>
<td>11.1</td>
</tr>
<tr>
<td>Prime-target delays</td>
<td>3.57</td>
<td>3.41</td>
</tr>
</tbody>
</table>

Table 1, target plausibility did not affect the proportion of direct hits, $F(1, 11) = .1$. There was a tendency to skip the plausible targets more often but it was not reliable, $F(1, 11) = 2.67, p = .13, MSE = .009$. In addition, a subjects x target plausibility repeated-measures ANOVA on delayed hits only showed no reliable effect on the number of fixations required to complete a gaze shift from prime to target, $F(1, 11) = .04$.

Prime fixations were analyzed in a subjects x target plausibility x gaze shift type (direct hit vs. skip vs. delayed hit) repeated-measures ANOVA. Table 2 shows that when an implausible target was wiggled extrafoveally the primes received reliably shorter first gazes ($F(1, 11) = 11.15, p < .007, MSE = 63,789$) and fewer fixations in the first gaze ($F(1, 11) = 11.89, p < .006, MSE = .386$). This effect was qualified by a target plausibility x gaze shift type interaction ($F(2, 19) = 6.14, p < .009, MSE = 34,085$, for gaze durations; $F(2, 19) = 3.05, p < .071, MSE = .342$, for first-pass refixations. Specifically, prime fixations preceding a direct gaze shift towards the target were unaffected by the target’s plausibility, while prime fixations preceding a delayed hit or a skip of the target were longer when the target was plausible.

While the more frequent and earlier fixation of implausible objects indicates that they are more salient saccade targets than plausible objects, there is no evidence that this is caused by a greater noticeability of the implausible-target wiggle which in turn is th
Table 2
Ongoing Prime Fixations as a Function of Wiggled Target’s Plausibility and Type of Subsequent Gaze Shift (SOA = 160 ms)

<table>
<thead>
<tr>
<th>Gaze shift type</th>
<th>Plausible target</th>
<th>Implausible target</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct hits</td>
<td>Skips</td>
</tr>
<tr>
<td>First fixation (ms)</td>
<td>392</td>
<td>351</td>
</tr>
<tr>
<td>First gaze (ms)</td>
<td>428</td>
<td>612</td>
</tr>
<tr>
<td>First-pass refixations</td>
<td>1.22</td>
<td>1.69</td>
</tr>
</tbody>
</table>

result of an immediate capture of attention. To the contrary, target plausibility had no effect on the frequency and speed of those gaze shifts which are most likely to be directly elicited by the wiggle, that is, the direct hits.

That target plausibility did affect prime fixations preceding a delayed target hit or a target skip is consistent with the notion that plausible targets are easier to process extrafoveally. The underlying rationale is that the latency of a gaze shift away from the prime is determined by the rate at which foveal prime and extrafoveal target information are acquired: As long as this rate does not drop below a minimum criterion, the gaze will remain on the prime. Because such a long prime gaze is partially caused by extensive extrafoveal target processing the need for foveal target analysis is reduced, resulting in a subsequent delay or even cancellation of target fixation.

Note that this argument is entirely consistent with explanations of similar data patterns in reading and word recognition. Specifically, Kennedy (this volume) reports that gaze durations on a foveal word are affected by length and initial trigram frequency of a simultaneously present parafoveal word. Kennedy interprets this as evidence for a model of eye-movement control in which the rate of foveal and parafoveal information intake is a determinant of when the eye will move. Further support for a relation between long gaze durations and extensive extrafoveal processing is found in reading where the skipping rate of words was found to increase when they were highly predictable from the preceding sentence (Balota, Rayner, & Pollatsek, 1985; Rayner & Well, 1996). This skipping effect was attributed to extrafoveal identification of predictable words because fixations on the preceding word were much longer, presumably due to the additional extrafoveal processing and the subsequent reprogramming of the next saccade to go beyond the identified word.

Sofar, the wiggle paradigm has yielded indications of a smaller useful field of view for implausible targets resulting in a greater need for foveal analysis. No evidence was found of a greater responsiveness to implausible-target wiggles caused by a stronger capture of attention. However, average speed and frequency of direct gaze shifts to the wiggling target may not be the best measure of wiggle responsiveness because direct hits can result from two categories of gaze shifts: Reactive shifts exogenously controlled by the wiggle, and active shifts endogenously controlled by the visual system's need for foveal analysis. Clearly, the former type would be most informative with respect to wiggle responsiveness: Objects that more frequently capture attention prior to the wiggle should elicit a greater proportion of truly reactive shifts towards the wiggle.
One way to distinguish between active and reactive gaze shifts is to look at time-locked effects of the wiggle on the distribution of single prime fixations. Previous research showed a disruptive effect of an abrupt visual onset on the duration of the ongoing fixation with a minimum delay of about 90 ms (Blanchard, McConkie, Zola, & Wolverton, 1984; McConkie, Underwood, Zola, & Wolverton, 1985) and a maximum impact after about 120 ms (van Diepen, De Graef, & d'Ydewalle, 1995). Given its constant timing, this effect on eye movements can be interpreted as a reflex-like orienting response to the onset. Assuming that the wiggle is a kind of abrupt visual onset, this means that the distribution of all single prime fixations of at least 160 ms should show a time-locked rise starting at wiggle SOA + 90 (= 250 ms) and peaking at wiggle SOA + 120 (= 280 ms). Such a peak in the distribution of gaze shift latencies, would indicate a sudden increase in the likelihood to reflexively shift away from the prime in response to the target wiggle. Figure 3 shows the distributions as a function of target plausibility. As expected, both distributions show a sudden rise in the 240-260 ms bin (i.e., wiggle SOA + 90 ms) which peaks in the 280-300 ms bin (i.e., wiggle SOA + 120 ms). The peak is somewhat larger for the implausible targets, which would be consistent with the notion of attentional capture by implausible objects. Much more striking, however, is the plausibility effect in the remainder of the distributions. While both distributions show a second peak between 340 and 400 ms, only the plausible-target distribution has a third peak between 420 and 460 ms. The multimodality of the distributions suggests that perhaps they are a multinomial mixture of three different underlying distributions, centered in the shaded areas in Figure 3. Yantis, Meyer, and Smith (1991) argue that a statistical test of this hypothesis requires pure samples of the underlying basis distributions, which in turn requires a theory of the different processes giving rise to the different distributions. Only then can one set up the experimental conditions under which pure samples of each distribution are most likely to be obtained. Obviously, this is beyond the scope of the present chapter which merely reanalyses earlier data. At this stage, I can only speculate about the origin of the observed distributions.

Basis for this speculation is an inspection of the locations targeted by the gaze shifts. Specifically, I first computed the distribution of direct prime-to-target shifts along the gaze shift latency continuum. As can be seen in Figure 4, direct hits (0 intermediate fixations) occurred most frequently following a gaze shift in the 340-400 ms region. A further delay of a direct gaze shift to the target was less likely if the target was implausible. For plausible targets, however, there was a second large concentration of direct hits following gaze shifts with latencies between 420-460 ms. Surprisingly, there were relatively few direct hits following gaze shifts in the 240-320 ms region which presumably includes reflexive orienting responses to the target wiggle. A visual inspection of scanpaths suggested, however, that this may be a matter of accuracy: Fast saccades away from the prime appeared to frequently just miss the target and were then followed by a quick corrective saccade landing on the target. When we include these ‘corrected’ hits (1 intermediate fixation), the likelihood of a reactive hit becomes comparable to that of later hits, demonstrating that the faster gaze shifts were indeed less accurate. Figure 5 (top panel) plots the resulting distributions of direct + corrected hits as a function of target plausibility and compares them with the distributions of gaze shifts to non-target locations (bottom panel).

The first thing to note is that the greater proportion of reflexive shifts in the overall implausible-target distribution of Figure 3 reflects more frequent gaze shifts to both target and non-target locations in that time band. This could mean that, regardless of where attention is allocated, peripheral motion of an implausible object is always a more noticeable event, perhaps because of its presumed featural dissimilarity from the scene (Rayner &
Pollatsek, 1992). Consequently, the distributions provide no strong evidence of meaning-based attentional capture by implausible objects, particularly when one takes into account the slightly greater density of plausible-target directed shifts at the very beginning of the reflexive time band (i.e., at wiggle SOA + 90 = 250 ms).

What the distributions do show is that gaze shifts in the 340-400 ms time band are preferentially directed towards the target: Regardless of target plausibility, the proportion of target-directed shifts peaks in this region, while that of non-target shifts remains constant or drops. Apparently, target saliency was at its maximum at this point and the timing coincides with that of voluntary saccades measured for stimuli identical to the line drawings used in the present experiments. Specifically, for viewers that were freely exploring these scenes for purposes of memorization or object search, Henderson et al. (1997) reported unimodal fixation duration distributions peaking at 210-220 ms. Thus, the 370 peak in the present distributions can be interpreted as a population of voluntary saccades towards objects that require further foveal analysis after their presence has been signaled by their motion 210 ms earlier (i.e., at wiggle SOA 160 ms).

Finally, for plausible targets only, gaze shifts to both targets and other locations show an outspoken frequency increase in the 420-460 ms range. The effect is strongest for shifts to non-target locations which show a steep recovery after a period of infrequency due to greater target saliency. Combined with the earlier finding of long fixation times preceding delayed target hits and target skips (Table 2) this suggests that the third peak is the product of a period of extensive extrafoveal attending to plausible targets, resulting in delayed saccades.

Summarizing the above, I would like to speculate that the distributions in Figure 3 are a mixture of three distinct populations of gaze shifts: 1) in the 240-320 ms region, involuntary, wiggle-evoked, and inaccurate gaze shifts; 2) in the 340-400 ms region, fast, voluntary, target-directed and accurate gaze shifts, prompted by an insufficient rate of extrafoveal target processing; 3) in the 420-460 ms region, slow, voluntary gaze shifts targeting or deliberately skipping peripherally located plausible target objects, delayed because of a high rate of context-supported extrafoveal target processing. As noted before, this hypothetical multinomial mixture of component distributions needs to be confirmed statistically in future research. Pertinent to the present discussion will be the question whether the three mixing proportions are reliably affected by the plausibility of the wiggled target. It seems safe to predict that, relative to plausible targets, implausible targets will prove to elicit more fast voluntary and fewer delayed voluntary gaze shifts, thus indicating that an implausible object constitutes a more salient target for a voluntary saccade because it is more difficult to process extrafoveally. Predictions are less clear with respect to context effects on the proportion of samples from the reflexive gaze shift population. On the one hand, the reflexive time band displayed a greater overall ocular reactivity to implausible-target wiggles. On the other hand, at the very onset of the reflexive band, gaze shifts specifically directed at the wiggling object were more likely when it was plausible. One way of reconciling these observations is to assume that attention was more likely to be captured by plausible targets at the time of the wiggle, but that the wiggle itself was a more noticeable and disruptive peripheral event when the wiggling object was implausible. The latter would be consistent with the notion of greater featural dissimilarity between implausible objects and their inconsistent environment, while rapid attentional capture by plausible objects would be in line with Christie and Klein's (1995) findings with an explicit motion detection paradigm.

To examine the viability of this interpretation, I analyzed data from a second wiggle study. Again, this study was conducted to examine context effects on the identification of fixated objects and thus provides no direct
Table 3
Effects of Wiggled Target’s Plausibility and Eccentricity on Direct Target Hits, Target Skips, and Number of Intervening Fixations preceding Delayed Target Hits (Wiggle SOA = 140 ms)

<table>
<thead>
<tr>
<th></th>
<th>Near Target</th>
<th></th>
<th>Far Target</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plausible</td>
<td>Implausible</td>
<td>Plausible</td>
<td>Implausible</td>
</tr>
<tr>
<td>Direct target hits (%)</td>
<td>62.6</td>
<td>64.6</td>
<td>49.5</td>
<td>60.3</td>
</tr>
<tr>
<td>Target skips (%)</td>
<td>11.2</td>
<td>4.0</td>
<td>6.6</td>
<td>4.1</td>
</tr>
<tr>
<td>Prime-target delays</td>
<td>4.01</td>
<td>2.84</td>
<td>3.07</td>
<td>1.81</td>
</tr>
</tbody>
</table>

tests of the hypothesised context effects on prefixational, attentional capture. However, the study did involve two changes from the first experiment which may be informative. First, if attention was indeed initially captured by plausible targets then this advantage should increase as the wiggle SOA is reduced from 160 ms to 140 ms. The capture effect Christie and Klein (1995) found for familiar targets was strong at 100 ms motion SOA but virtually disappeared by 200 ms, so the 160 ms SOA wiggle in the first study may only have revealed the tail of the plausible capture effect.

Second, if implausible-target motion is more noticeable in peripheral vision then this advantage should be reduced by bringing the moving object closer to the fixation point. In this second study, we systematically manipulated the distance between the wiggled target and the preceding prime. Each of 16 targets could appear in a 6 s-exposure of a plausible or implausible scene, but orthogonal to this they could also be near or far to the point of fixation when they wiggled. This manipulation of target eccentricity was achieved by simply creating two versions of each plausible and implausible scene: one with the target at about 3° from the prime and one with the target at about 8° from the prime. A targets x target plausibility x target eccentricity analysis of prime-target distances revealed \( F(1, 15) < 1 \) for all effects involving target plausibility. Eye-movement data were collected from 12 new participants and with the exceptions noted above experimental procedure and stimuli were identical to that of the first study.

Table 3 presents the means obtained in a subjects x target plausibility x target eccentricity repeated-measures ANOVA of all trials on which the wiggle occurred during the first scene fixation. The proportion of trials on which the eye saccaded directly to the wiggled target was affected by target plausibility and target eccentricity: Implausible targets produced more captures \( (F(1, 11) = 6.74, p < .03, \text{MSE} = .007) \), and so did near targets \( (F(1, 11) = 11.25, p < .007, \text{MSE} = .008) \). The plausibility effect appeared to be stronger for far targets, but this interaction was not reliable \( (F(1, 11) = 1.2, p < .3, \text{MSE} = .019) \). Plausible targets were also marginally more likely to be left unfixated \( (F(1, 11) = 4.41, p < .06, \text{MSE} = .006) \) an effect which appeared to be larger for near targets but this interaction was not reliable \( (F(1, 11) = 2.1, p < .18, \text{MSE} = .003) \). In addition, a subjects x target plausibility x target eccentricity repeated-measures ANOVA on delayed target hits revealed that it took more fixations to complete the prime-to-target gaze shift when the target was plausible \( (F(1, 11) = 4.83, p < .05, \text{MSE} = 11.38) \) or when the target was near to the initial scene fixation \( (F(1, 11) = 4.15, p < .06, \text{MSE} = 8.64) \). All these data are consistent with the interpretation that the need for foveal analysis of a given object determines
Table 4
Ongoing Prime Fixations as a Function of Wiggled Target’s Plausibility, Target Eccentricity, and Destination of Gaze Shift (SOA = 140 ms)

<table>
<thead>
<tr>
<th>Gaze shift destination</th>
<th>Plausible target</th>
<th>Implausible target</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target</td>
<td>Other</td>
</tr>
<tr>
<td>Near targets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First fixation (ms)</td>
<td>340</td>
<td>297</td>
</tr>
<tr>
<td>First gaze (ms)</td>
<td>378</td>
<td>343</td>
</tr>
<tr>
<td>First-pass refixations</td>
<td>1.17</td>
<td>1.16</td>
</tr>
<tr>
<td>Far targets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First fixation (ms)</td>
<td>356</td>
<td>341</td>
</tr>
<tr>
<td>First gaze (ms)</td>
<td>377</td>
<td>406</td>
</tr>
<tr>
<td>First-pass refixations</td>
<td>1.14</td>
<td>1.25</td>
</tr>
</tbody>
</table>

its saliency as a saccade target, and that the need for foveal analysis is greater for implausible and more eccentric objects.

Prime fixations that were ongoing at the time of the wiggle were entered in a subjects x target plausibility x target eccentricity x gaze shift destination (target vs. other) ANOVA. The latter variable divides the observations as a function of where the eye shifted after leaving the prime: Target-directed shifts with maximum 1 fixation in the prime-target interval, and shifts directed away from the target. The analysis showed no reliable effects involving target plausibility, although Table 4 reveals that the pattern for far targets was identical to that observed in Table 2: Prime fixations preceding a target-directed gaze shift were unaffected by the target’s plausibility, while prime fixations preceding a delayed hit or a skip of the target were longer when the target was plausible. Earlier, I interpreted this pattern as suggesting context-supported extended extrafoveal processing of plausible objects, delaying or cancelling the need for a saccade towards these objects. That the pattern was absent for near targets is consistent with the idea that context provides little additional benefit when extrafoveal preview quality is high (Henderson, 1992b).

J udging from a comparison of Tables 1 and 3, reducing the wiggle SOA by 20 ms did not decrease but increased the greater saliency of implausible targets as reflected in the higher frequency and shorter delays with which they were targeted by saccades. Based on Table 2, I argued that this saliency does not result from faster attentional capture by implausible objects but from a greater need for foveal analysis and this interpretation is again confirmed the similar pattern in Table 4. When one looks at the distributions of gaze shift latencies in this second study, this conclusion is corroborated further.

Figure 6 plots gaze shift distributions as a function of target plausibility and eccentricity. The top panels are based on all gaze shifts, the bottom panels exclude gaze shifts that were not directed at the target but these only amounted to 20% of the data which explains the similarity between top and bottom panels. For far targets,
the data replicate the findings from the first study (Figure 3): Three gaze shift populations for plausible targets and only two for implausible targets where the delayed-voluntary peak is virtually absent. Note, however, that, with the exception of the fast-voluntary distributions at 370 ms, the timing of the peaks has changed. For plausible targets, the reflexive peak shifted forward by 20 ms as was to be expected given the reduction of the wiggle SOA by 20 ms. In contrast, for implausible targets the reflexive peak shifted backward and moved outside of the reflexive time band (i.e., wiggle SOA + 120 ± 30 ms). Finally, the delayed voluntary peak also shifted forward by 20 ms, moving outside of the time band hypothesised from the first study. This could indicate that, like the reflexive gaze shifts, the delayed voluntary shifts are also time-locked. Specifically, one could argue that if attention is summoned 20 ms earlier by the wiggle, then extrafoveal target processing can begin 20 ms earlier and will end 20 ms earlier. Following this rationale, the constant, fast-voluntary shifts at 370 ms aren’t locked to the wiggle which runs counter to the earlier suggestion that they follow the wiggle onset with a saccadic latency that is modal for free exploration of the particular stimuli used in these studies. Obviously, a definitive settlement of these timing issues will require more research using a control distribution of gaze shifts in the absence of a wiggle. It is quite clear, however, that speeding up the wiggle resulted in a separation of the reflexive distributions for plausible and implausible targets. Gaze shifts to plausible targets speeded up in synchrony with the reduction in wiggle SOA supporting the hypothesis that attention was already captured by the plausible target at the onset of its wiggle. Gaze shifts to implausible targets slowed down, suggesting that by speeding it up the wiggle now preceded the allocation of attention to the target. Thus, the data indicate that by reducing the wiggle SOA, it now occurs within an initial time interval during which plausible objects have attentional precedence over implausible objects.

This interpretation is strengthened by inspection of the near-target distributions where plausible targets display a similar advantage. Also note that the delayed-voluntary peak in these distributions is considerably smaller than for the far targets. This probably reflects the fact that identification of near targets in parafoveal vision requires much less surplus processing time than target identification in peripheral vision. Finally, the comparison of near and far targets shows that the slowed, reflexive peak for the implausible targets is much smaller for the near targets. This is consistent with the hypothesis that a greater overall reactivity to implausible-target wiggles reflects a greater peripheral noticeability based on featural target-background dissimilarity.

Conclusions

When research on object perception in real-world scenes is referenced in other domains, this is usually done to illustrate the existence of semantic context effects on stimulus identification and on stimulus selection. I started this chapter by pointing out that scene researchers agree on the first effect (although see Henderson & Hollingworth, 1997 for a recent vote of disagreement), but are not convinced with respect to the second. Some authors see no role for context as a determinant of where attention and the eye will go, and others disagree as to what the direction of the effect is: Semantically plausible objects might be detected across a greater extent of the visual field and thus receive processing preference, or alternatively, semantically implausible objects might pop out of the scene and therefore attract attention.

While scene research had a long standing theory on why plausible objects should be more perceptible in peripheral vision (i.e., schema-driven perception), no such explanation existed for implausible popout which appeared to be supported exclusively by a single, influential observation of a fixation precedence for
objects that do not belong in a scene. Nevertheless, implausible popout was not classified as a U.F.O. (Unreproducible Freak Observation). This should probably be attributed to the intuition that implausible popout would be good from an evolutionary perspective: Slowed detection of the unexpected tiger on the kitchen counter would not contribute to the success of the human species. In support of implausible popout, however, Johnston and colleagues recently outlined and tested a computational model of visual input processing in which competing influences of schema-driven perception and novel popout are detailed and reconciled. Based on this theory, scene research can now test for specific patterns of effects as well as processing characteristics which would be diagnostic for schema-driven perception and/or implausible popout.

In a first attempt to contribute to this enterprise, I presented some data relevant to the hypothesis that implausible objects are more likely than plausible objects to capture the processing focus early on in scene exploration. If this were true, features or events at an implausible target location should be detected more quickly. Specifically, rapid and repeated movement of a peripherally located object should elicit faster and/or more frequent gaze shifts if the wiggled object was implausible and therefore more likely to be at the center of ongoing data-driven processing. While objects that did not belong, proved to be more salient saccade targets, a detailed analysis of gaze shift latencies and destinations revealed that after 160 ms of scene processing there was no unequivocal processing precedence for these objects: Plausible-target wiggles also elicited reflexive orienting responses and the saliency of implausible objects appeared to be caused by a greater difficulty to process them in extrafoveal vision.

It is not clear at this point how the larger useful field of view for plausible objects should be understood. One possibility is that top-down activation from context combines with bottom-up activation from extrafoveal processing and thus increases the diagnosticity of extrafoveal features because context has narrowed down the possible identities of the extrafoveal target. This seems to be the view held in explanations of context effects on word skipping in reading (e.g., Rayner & Morris, 1992; Rayner & Well, 1996). Another possibility is that contextual activation directly enhances the quantity and speed at which extrafoveal target features are encoded from plausible objects in the image. Consistent with this notion is the claim that conceptually activated objects exert a very early pull on attention (e.g., Christie & Klein, 1995; Hoffmann, 1987) thus enhancing perceptual sensitivity at the attended location (e.g., Downing, 1988; Hawkins, Hillyard, Luck, Mouloua, Downing, & Woodward, 1990). This claim was supported in a between-experiment comparison of wiggle detectability at 160 and 140 ms SOA. At the shorter SOA, plausible-target wiggles were more likely than implausible-target wiggles to elicit reflexive orienting responses and the effect decreased at the longer SOA.

Naturally, a within-experiment parametric variation of wiggle SOA is needed to unambiguously establish the suggested interaction between processing interval and processing precedence. Given the fine-grained analysis that appears to be required to detect the signs of processing precedence, this does not promise to be an easy task. However, if we want to improve our understanding of the blend of concept-driven and data-driven processing in real-world scene perception we will have to allow for the possibility that the exact dosages may vary over time, both within a single fixation and across the whole extent of scene exploration.

References


Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding


Kennedy, A. (this volume). *Paraofoveal-on-foveal effects in word identification? Evidence for a processing trade-off?*


Author Notes

The writing of this chapter was supported by GOA-grant 93/1 from the Research Council of the University of Leuven, and by research grant G.2058.94 from the Fund for Scientific Research-Flanders (F.W.O.-Vlaanderen). I also want to thank Géry d’Ydewalle, John Findlay and an anonymous reviewer for their helpful comments on an earlier draft.
Figure 1. Course of a trial in the wiggle paradigm. Following a fixation of minimum 200 ms on a fixation cross, an 8-s scene exposure is automatically initiated, resulting in fixation on the prime object (circled for expository purposes). After a 160 ms delay, a 120 ms wiggle is started in which a designated target object moves up and down twice with an amplitude of 4 min of arc (arrow illustrates motion). The wiggle is intended to elicit a saccade from prime to wiggled target (example of ideal saccade superimposed on scene).
Figure 2. Example of plausible (left panel) and implausible (right panel) conditions for the target rolling pin. Scene exploration for these stimuli would start at the computer in the implausible office background, and at the blender in the plausible kitchen background.
Figure 3. Relative frequency distribution of gaze shift latencies away from the prime as a function of the wiggled target’s plausibility. All distributions are based on single-fixation prime gazes of at least 160 ms. Bin-size is 20 ms and the graphs plot the midpoints of the bins. Dark bands indicate hypothetical location for reflexive (240-320), fast voluntary (340-400), and delayed voluntary (420-460) gaze shift distributions.
**Figure 4.** Relative frequency distribution of latencies to shift gaze from prime to target. *Interval 0* distributions plot direct prime-to-target shifts as a function of target plausibility, *interval 0+1* distributions do the same for prime-to-target shifts with maximum 1 intervening fixation. Construction of distributions and time bands is identical to Figure 3.
Figure 5. Relative frequency distribution of gaze shift latencies away from the prime as a function of the wiggled target’s plausibility and the destination of the gaze shift. Top panel plots gaze shifts that landed on the target after maximum 1 intervening fixation, bottom panel plots all other gaze shifts. Construction of distributions and time bands is identical to Figure 3.
Figure 6. Relative frequency distribution of gaze shift latencies away from the prime as a function of the wiggled target’s plausibility and eccentricity. All distributions are based on single-fixation prime gazes of at least 140 ms. Bin-size is 20 ms and the graphs plot the midpoints of the bins. Dark bands indicate hypothetical location for reflexive (220-300), fast voluntary (340-400), and delayed voluntary (420-460) gaze shift distributions. Top panels plot all gaze shifts regardless of destination, bottom panels only plot gaze shifts that landed on the target after maximum 1 intervening fixation.